

## Description

# METHOD AND STRUCTURE FOR FORMING STRAINED Si FOR CMOS DEVICES

### BACKGROUND OF INVENTION

[0001] The invention generally relates to methods for manufacturing a semiconductor device with improved device performance, and more particularly to methods for manufacturing semiconductor devices which impose tensile and compressive stresses in the substrate of the device during device fabrication.

[0002] Generally, metal-oxide semiconductor transistors include a substrate made of a semiconductor material, such as silicon. The transistors typically include a source region, a channel region and a drain region within the substrate. The channel region is located between the source and the drain regions. A gate stack, which usually includes a conductive material, a gate oxide layer and sidewall spacers, is generally provided above the channel region. More par-

ticularly, the gate oxide layer is typically provided on the substrate over the channel region, while the gate conductor is usually provided above the gate oxide layer. The sidewall spacers help protect the sidewalls of the gate conductor.

[0003] It is known that the amount of current flowing through a channel which has a given electric field across it, is generally directly proportional to the mobility of the carriers in the channel. Thus, by increasing the mobility of the carriers in the channel, the operation speed of the transistor can be increased.

[0004] It is further known that mechanical stresses within a semiconductor device substrate can modulate device performance by, for example, increasing the mobility of the carriers in the semiconductor device. That is, stresses within a semiconductor device are known to enhance semiconductor device characteristics. Thus, to improve the characteristics of a semiconductor device, tensile and/or compressive stresses are created in the channel of the n-type devices (e.g., NFETs) and/or p-type devices (e.g., PFETs). However, the same stress component, for example tensile stress or compressive stress, improves the device characteristics of one type of device (i.e., n-type device or

p-type device) while discriminatively affecting the characteristics of the other type device.

[0005] In order to maximize the performance of both NFETs and PFETs within integrated circuit (IC) devices, the stress components should be engineered and applied differently for NFETs and PFETs. That is, because the type of stress which is beneficial for the performance of an NFET is generally disadvantageous for the performance of the PFET. More particularly, when a device is in tension (in the direction of current flow in a planar device), the performance characteristics of the NFET are enhanced while the performance characteristics of the PFET are diminished. To selectively create tensile stress in an NFET and compressive stress in a PFET, distinctive processes and different combinations of materials are used.

[0006] For example, a trench isolation structure has been proposed for forming the appropriate stresses in the NFETs and PFETs, respectively. When this method is used, the isolation region for the NFET device contains a first isolation material which applies a first type of mechanical stress on the NFET device in a longitudinal direction (parallel to the direction of current flow) and in a transverse direction (perpendicular to the direction of current

flow). Further, a first isolation region and a second isolation region are provided for the PFET and each of the isolation regions of the PFET device applies a unique mechanical stress on the PFET device in the transverse and longitudinal directions.

[0007] Alternatively, liners on gate sidewalls have been proposed to selectively induce the appropriate strain in the channels of the FET devices (see Ootsuka et al., IEDM 2000, p. 575, for example). By providing liners the appropriate stress is applied closer to the device than the stress applied as a result of the trench isolation fill technique.

[0008] While these methods do provide structures that have tensile stresses being applied to the NFET device and the compressive stresses being applied along the longitudinal direction of the PFET device, they may require additional materials and/or more complex processing, and thus, resulting in higher cost. Further, the level of stress that can be applied in these situations is typically moderate (i.e., on the order of 100s of MPa). Thus, it is desired to provide more cost-effective and simplified methods for creating large tensile and compressive stresses in the channels NFETs and PFETs, respectively.

## **SUMMARY OF INVENTION**

[0009] In a first aspect of the invention, the invention provides a method for manufacturing a device including an n-type device and a p-type device. The method involves doping a portion of a semiconductor substrate and forming a gap in the semiconductor substrate by removing at least a portion of the doped portion of the semiconductor substrate. The method further involves growing a strain layer in at least a portion of the gap in the semiconductor substrate.

[0010] In aspects of the invention, for the n-type device, the strain layer is grown on at least a portion which is substantially directly under a channel of the n-type device. For the p-type device, the strain layer is grown on at least a portion which is substantially directly under a source region or drain region of the p-type device and not substantially under a channel of the p-type device.

[0011] In another aspect of the invention, the invention provides a method for manufacturing a device including an n-type device and a p-type device. The method involves growing a strain layer on a semiconductor substrate and growing a silicon layer above the strain layer. A gap is formed between the semiconductor substrate and the silicon layer by removing at least a portion of the silicon layer and the

strain layer from above the semiconductor substrate and a strain layer is grown in at least a portion of the gap. For the n-type device, the strain layer is grown on at least a portion which is substantially directly under a channel of the n-type device. For the p-type device, the strain layer is grown on at least a portion which is substantially directly under a source region or drain region of the p-type device and not substantially under a channel of the p-type device.

[0012] This invention separately provides a semiconductor device which has a semiconductor substrate having at least one gap, the gap extending under a portion of the semiconductor substrate. The device includes a gate stack on the semiconductor substrate and a strain layer formed in at least a portion of the gap, where the gap is formed by doping a portion of the semiconductor substrate and etching the doped portions of the semiconductor substrate.

[0013] In another aspect of the invention, the invention provides a semiconductor device which has a semiconductor substrate having at least one gap, the gap extending under a portion of the semiconductor substrate. The device includes a gate stack on the semiconductor substrate and a

strain layer formed only under at least a portion of a source region and a drain region of the semiconductor substrate.

[0014]

## **BRIEF DESCRIPTION OF DRAWINGS**

[0015] Figure 1 illustrates desired stress states for PFETs and NFETs;

[0016] Figures 2(a) through 2(j) illustrate an exemplary process for forming a p-type transistor according to the invention;

[0017] Figures 3(a) through 3(d) illustrate an exemplary process for forming an n-type transistor according to the invention;

[0018] Figure 4 illustrates a top-down view of a transistor according to the invention; and

[0019] Figure 5 shows a cross-section of a semiconductor substrate according to the invention using a scanning electron microscope.

## **DETAILED DESCRIPTION**

[0020] The invention provides a method for fabricating devices with improved performance characteristics. When a stress layer, such as, a SiGe layer, a  $\text{Si}_3\text{N}_4$  layer, a  $\text{SiO}_2$  layer or a  $\text{SiO}_x\text{N}_y$  layer is grown epitaxially on a silicon layer, com-

pressive forces form within the SiGe layer and tensile forces form in the silicon layer. In an aspect of the invention, the silicon substrate has a gap in which a strain layer is grown. The gap includes a tunnel-like portion which is between an upper portion of the semiconductor substrate and a lower portion of the semiconductor substrate. More particularly, the upper portion has a lower surface and the lower portion face has an upper surface and the lower surface of the upper portion faces the upper surface of the lower portion. By having a strain layer substantially below a channel and/or a strain layer in a region of the semiconductor substrate substantially below a source region and/or a drain region of the semiconductor device stresses are formed in the channel of the transistor. In an aspect of the invention, the gap in the silicon substrate is formed by selectively etching the silicon substrate and then epitaxially growing SiGe on the silicon substrate.

[0021] Tensile and/or compressive stresses can be provided in the channel of a transistor depending on the proximity of the grown SiGe to the channel of the transistor. By selectively etching the silicon layer below a transistor and selectively growing SiGe on the etched portion of the silicon layer, tensile stresses can be provided in the channel of



NFETs and compressive stresses can be provided in the channel of PFETs. Further, by implementing the stresses by selectively etching a portion of the silicon below a transistor prior to growing SiGe, this invention provides stress levels in the silicon under the gate (e.g., the channel region) which are much larger than the isolation-based or liner-based approaches.

[0022] In this invention, a stress layer, such as a SiGe layer, for example, is used to form stresses in a channel of the semiconductor device. When a SiGe layer is grown on a semiconductor layer the surrounding semiconductor material is subjected to tensile stress while the grown SiGe layer is subjected to compressive stress. In particular, a portion of the semiconductor device is put under tensile stress and the SiGe layer is subjected to compressive stress because the SiGe layer has a different lattice structure than the silicon layer. Further, the stress levels resulting from the SiGe stress layer are relatively high (on the order of 1–2 GPa).

[0023] However, as discussed above, tensile stresses in the channel area are beneficial to the NFET drive currents while compressive stresses in the channel area are beneficial to the PFET drive currents. In particular, tensile stresses sig-

nificantly hinder the PFET drive currents. In this invention, the stresses in the PFET are made to be compressive stresses rather than tensile stresses in order to improve the performance of the PFET. Thus, this invention provides a method for providing longitudinal compressive stresses along the channel of the PFET while providing tensile stresses along the channel of the NFET to improve the performance of the devices.

[0024] Figure 1 illustrates desired stress states for improving the performance of PFETs and NFETs (see Wang et al., IEEE Tran. Electron Dev., v.50, p.529, 2003). In Figure 1, an NFET and a PFET are shown to have a source region, a gate region and a drain region. The NFET and PFET are shown to have arrows extending outward from the active area to illustrate tensile stresses. The arrows extending inward toward the PFET device are illustrative of compressive forces. More specifically, the outwardly extending arrows, shown extending from the NFET, illustrate a tensile stress that is desired in the transverse and longitudinal directions of the device. On the other hand, the inwardly extending arrows, shown with relation to the PFET, illustrate a desired longitudinal compressive stress.

[0025] The range of stresses needed to influence device drive

currents is of the order of a few hundred MPa to a few GPa. The width and the length of the active area of each device are represented by "W" and "L", respectively. It should be understood that each of the longitudinal or transverse stress components could be individually tailored to provide the performance enhancements for both devices (i.e., the NFET and the PFET).

[0026] Figures 2(a) through 2(j) depict an exemplary process for forming n-type devices according to this invention. As shown in Figure 2(a), a patterned photo-resist layer 205 is deposited over a silicon substrate 200 and the exposed portion of the silicon substrate 200 is doped, for example, with Ge, As, B, In or Sb. For example, the doping concentration of Ge may be, for example, about  $1 \times 10^{14}$  Ge/cm<sup>2</sup> to about  $1 \times 10^{16}$  Ge/cm<sup>2</sup>. A doped region 207 is formed in the semiconductor substrate 200.

[0027] Then, as shown in Figure 2(b), the patterned photo-resist layer 205 is removed and a mask 210, made of nitride, for example, is deposited on the surface of the semiconductor substrate 200. The mask 210 protects the semiconductor substrate beneath it from being etched during reactive ion etching (RIE). Generally, the mask 210 exposes portions of the semiconductor substrate where shallow

trenches are to be formed via RIE.

[0028] As shown in Figure 2(c), RIE is performed to form grooves/trenches 215 in the semiconductor substrate 200. As a result of the RIE step, side-wall portions 217 of the doped semiconductor region are formed. In particular, the location of the formed grooves/trenches at least partially overlaps a portion of the doped semiconductor region 207 such that when the grooves/trenches 215 are formed, the doped semiconductor substrate region is exposed. Further, as will be discussed below, after a strain layer is formed, oxide material is deposited to fill the trenches, such that devices adjacent to each other on the semiconductor substrate 200 are electrically isolated from each other.

[0029] After the grooves/trenches 215 are formed, wet etching and/or dry etching is performed to selectively remove the doped semiconductor 207. Generally, the depth of the trench will be about 1000 Angstroms to about 5000 Angstroms from the upper surface 231 (Figure 2(f)) of the semiconductor substrate and the thickness of a channel region of a transistor is typically about 30 Angstroms to about 200 Angstroms.

[0030] As shown in Figure 2(d), etching may be performed until a

tunnel-like gap 219 is formed between an upper portion 221 of the semiconductor substrate 200 and a lower portion 223 of the semiconductor substrate 200. Typically, a portion having a depth of about 300 Angstroms to about 5000 Angstroms is etched from the semiconductor substrate 200. In the case of an n-type transistor it is desired to form the strain layer substantially directly and/or directly under the channel of the device. Thus, for n-type transistors there is at least a gap under the channel of the device.

[0031] Next, as shown in Figure 2(e), a spacer material 225 is deposited over the semiconductor substrate 200. The spacer material may be, for example, a non-conformal film such as, silicon carbide SiC, oxynitride or a film stack, such as, an oxide film and a nitride film. This spacer material 225 is formed on the exposed portions of the semiconductor substrate 200 other than the portion of the semiconductor substrate below the upper portion 221.

[0032] As shown in Figure 2(f), a strain layer 227 is epitaxially grown in the tunnel-like gap 219 of the semiconductor substrate 200. As shown in Figure 2(f), the strain layer 227 is generally formed between an upper portion 221 and a lower portion 223 of the semiconductor substrate

200, where the upper portion 221 of the semiconductor substrate 200 is part of the original semiconductor substrate (i.e., was not removed/disturbed and deposited). That is, the strain layer 227 is generally formed via selective deposition such that the strain layer 227 is formed on the exposed surfaces of the semiconductor substrate 200. Further, because the strain layer 227 is formed in a tunnel-like gap, the upper surface 231 of the upper portion 221 is undisturbed (i.e., original) and substantially flat.

[0033] The strain layer may be, for example, silicon germanium or silicon carbide. It should be understood that the strain layer may be made of any known appropriate material.

[0034] After the strain layer 227 is formed, the spacer material 225 is removed using wet chemicals. It should be understood that any known applicable method may be used to remove the spacer material 225. The resulting device without the spacer material is shown in Figure 2(g).

[0035] As discussed above, and as shown in Figure 2(h), oxide material 233 is then deposited to fill the trenches and electrically isolate the device from any adjacent device. After filling the trenches with the oxide material, the mask 210 is removed using any known appropriate method. After the mask 210 is removed, chemical mechanical polish-

ing (CMP) is performed to substantially planarize the upper surface 231 of the semiconductor substrate 200.

[0036] Next, the semiconductor device is further fabricated using known methods. For example, as shown in Figure 2(i), a gate oxide layer 235 is grown on the upper surface 231 of the semiconductor substrate 200. A gate oxide layer 235 of about 10 Angstroms to about 100Å is generally grown. On the gate oxide layer 235, a polysilicon layer 236 is generally deposited using chemical vapor deposition (CVD) to a thickness of about 500 Angstroms to about 1500 Angstroms to form the gate electrode 237. Patterned photoresist layers (not shown) are used to define the gate electrodes. A thin layer of oxide (not oxide) is then grown on the remaining polysilicon. Patterned photoresist layers (not shown), which are later removed, are used to successively tip (and halo counteracting doping implants) implant the n-type and p-type transistors. For n-type transistors, a very shallow and low dose implant of arsenic ions, for example, may be used to form the p-tip (while a Boron implant, for example, may be used for halos). For p-type transistors, (discussed below with regards to Figures 3(a) 3(d)), a very shallow and low dose implant of  $\text{BF}_2$  ions, for example, may be used to form n-tip (while an arsenic im-

plant may, for example, be used for halos).

[0037] Next, as shown in Figure 2(j), spacers 238 may be formed by depositing a silicon nitride layer (not shown) using CVD to a thickness of about 100 Angstroms to about 1000 Angstroms and then etching the nitride from the regions other than the sidewalls of the gate. The combination of the gate oxide layer 235, gate electrode 237 and spacers 238 may be referred to as a gate stack.

[0038] Patterned photoresist layers (not shown), which are removed prior to the next stage of the process, are used to successively create the source/drain regions of the transistors. For the n-type transistors, a shallow and high-dose of arsenic ions, for example, may be used to form the source/drain regions 240 and 241 while the p-type transistors are covered with the corresponding photoresist layer. As discussed above, in the methods according to this invention, the source and drain regions 240 and 241 are formed in upper portions of semiconductor substrate 200 (i.e., not removed and reformed). For the p-type transistors, (discussed below with regards to Figures 3(a) 3(d)), a shallow and high dose of  $\text{BF}_2$  ions, for example, may be used to form the source/drain regions 30 while the n-type transistors are covered with the corresponding



photoresist layer. An anneal is then used to activate the implants. The exposed oxide on the structure is then stripped by dipping the structure in HF in order to expose bare silicon in the source, gate and drain regions of the transistors.

[0039] Still referring to Figure 2(j), metal is deposited to a thickness of about 30 Angstroms to about 200 Angstroms across the wafer surface in order to form silicide 242. The silicide could be formed from reacting the underlying with any deposited metal such as Co, Hf, Mo, Ni, Pd<sub>2</sub>, Pt, Ta, Ti, W, and Zr. In the regions, such as, the source, drain and gate regions, where the deposited metal is in contact with silicon, the deposited metal reacts with the silicon to form silicide. Next, the structure is heated to temperature of about 300°C to about 1000°C to allow the deposited silicide material to react with the exposed polysilicon or silicon. During sintering, silicide only forms in the regions where metal is in direct contact with silicon or polysilicon. In the other regions (i.e., where the deposited metal is not in contact with silicon), the deposited metal remains unchanged. This process aligns the silicide to the exposed silicon and is called "self-aligned silicide" or salicide. The unreacted metal is then removed using a wet etch while

the formed silicide remains.

[0040] In the methods according to this invention because the source and drain regions of the semiconductor device are formed on portions of the semiconductor substrate which are undisturbed (i.e., not etched and re-formed), the surface is more favorable to cobalt silicide formation as cobalt silicide. Further, generally an oxide fill (not shown) followed by chemical mechanical polishing is used to planarize the surface. The fabrication processes continues as necessary according to the design specifications.

[0041] Figures 3(a) through 3(d) depict an exemplary process for forming p-type devices according to this invention. The process for forming p-type devices is similar to the process for forming n-type devices, as discussed above with regards to Figures 2(a) 2(j) and thus, the following discussion will primarily focus on the differences between the two process. The details of the method for forming a p-type device which are not discussed below, may be found in the above description of the method for forming an n-type device.

[0042] As shown in Figure 3(a), a patterned photo-resist layer 305 is deposited. For the p-type devices, the portion 307 of the semiconductor substrate 300 which will be below

the channel of the semiconductor device is also covered with the patterned photo-resist layer 305. Thus, for the p-type devices, as shown in Figure 3(b), when the doped region of the semiconductor substrate is selectively etched to form the gap 315, a portion 308 of the semiconductor substrate 300 remains. After the structure is formed, this portion 308 of the semiconductor substrate is substantially directly under the channel of the semiconductor device.

[0043] Next, as shown in Figure 3(c), a strain layer 327 is grown in the gap between the remaining upper portion 301 and lower portion 302 of the semiconductor substrate 300. Then, as shown in Figure 3(d), oxide material is deposited to fill the gaps/trenches 315. Similar to the process for forming n-type devices, gate oxide 335 is deposited on the upper surface of the semiconductor substrate and the gate electrode 337, spacers 338, source/drain regions 340 and 341 and silicide contacts 342 are formed.

[0044] Figure 4 depicts a top-down view of a transistor according to the invention. A cross-sectional view taken along line A-A of Figure 4 is the structure shown in Figure 2(i) and a cross-sectional view taken along line B-B of Figure 4 is the structure shown in Figure 2(j). As shown in Figure 4,

the gate electrode 242 with the spacer 238 is located above the semiconductor substrate 200. The oxide fill 233 (i.e., shallow trench isolation structure) isolates the source and drain regions 240 and 241 of the semiconductor substrate 200.

[0045] Figure 5 shows a cross-section of a semiconductor substrate according to the invention. The representation of the semiconductor substrate shown in Figure 5 was obtained using a scanning electron microscope. In particular, Figure 5 shows the silicon substrate after the doped silicon has been selectively removed to form tunnel-like gaps 219 in the semiconductor substrate. As shown in Figure 5, a lower surface of an upper portion of the semiconductor substrate and an upper surface of a lower portion of the semiconductor substrate define a portion of the gap in the semiconductor substrate. The gap in the semiconductor substrate may include an opening along an upper surface of the semiconductor substrate.

[0046] In another embodiment of the methods according to this invention, instead of selectively doping the semiconductor substrate with Ge, for example, such that selective portions of the semiconductor substrate may be removed via etching, it is possible to grow a layer, such as, a SiGe

layer, on the semiconductor substrate, followed by a silicon epitaxial layer, for example. Then, similar to the doping method described above, sidewalls of the SiGe may be exposed and then selectively etched to form the gaps in the semiconductor substrate.

[0047] As discussed above with regards to Figure 1, in PFETs, a longitudinal compressive stress is desired. The typical range for the desired compressive/tensile stresses is on the order of a few hundred MPa to a few GPa. For example, stresses of about 100MPa to about 2 or 3 GPa are generally desired. The invention can produce very high compressive stresses and tensile stresses in the channels of the PFET and NFET devices, respectively.

[0048] By providing tensile stresses to the channel of the NFET and compressive stresses to the channel of the PFET the charge mobility along the channels of each device is enhanced. Thus, as described above, the invention provides a method for providing compressive stresses along the longitudinal direction of the channel by providing a strain layer either substantially directly under the channel of the semiconductor device or substantially directly under the source and/or drain region of the semiconductor device. This invention also provides a method for optimizing the

stress level in the transistor channel by adjusting the location and/depth of the gap where the strain layer is formed.

[0049] While the invention has been described in terms of embodiments, those skilled in the art will recognize that the invention can be practiced with modification within the spirit and scope of the appended claims.